Thoughts on Verification at ISI Timescales

Barbara Brown

Joint Numerical Testbed Program
National Center for Atmospheric Research
Boulder CO

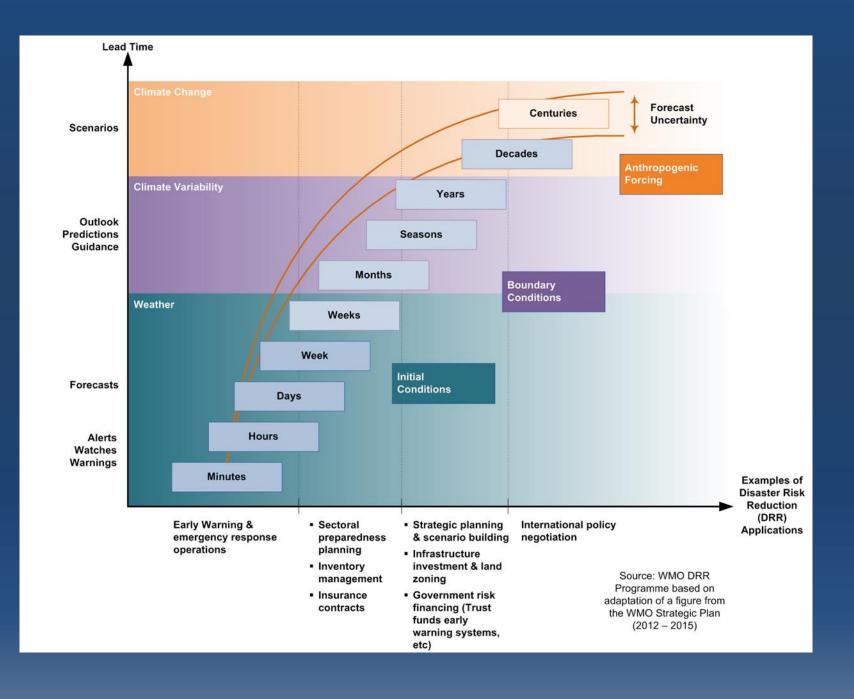
ESPC & NUOPC Workshop

on

Construction, Visualization and Verification of Ensemble-based Predictions from Sub-seasonal to ISI Timescales

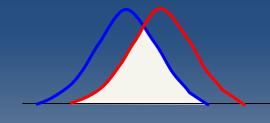
Seamless verification

Seamless forecasts consistent across space/time scales single modelling system or blended likely to be probabilistic / ensemble climate seasonal global prediction prediction change seasonal Spatial scale NWP prediction very regional short range local nowcasts point minutes hours weeks months decades days years forecast aggregation time



"Seamless verification" – consistent across space/time scales

- Modelling perspective is my model doing the right thing?
 - Process approaches
 - LES-style verification of NWP runs (first few hours)
 - T-AMIP style verification of coupled / climate runs (first few days)
 - Single column model
- Statistical approaches
 - Spatial and temporal spectra
 - Spread-skill
 - Marginal distributions (histograms, etc.)



"Seamless verification" – consistent across space/time scales

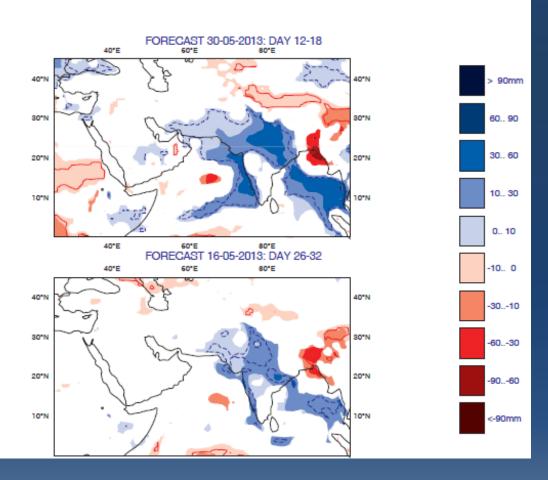
- User perspective can I use this forecast to help me make a better decision?
 - Neighborhood approaches spatial and temporal scales with useful skill
 - Object-based approaches what are the errors in attributes that matter (e.g., spatial displacements)
 - Generalized discrimination score (Mason & Weigel, MWR 2009)
 - consistent treatment of binary, multi-category, continuous, probabilistic forecasts
 - Calibration accounting for space-time dependence of bias and accuracy?
 - Conditional verification based on larger scale regime
 - Extreme Forecast Index (EFI) approach for extremes

ANALYSIS 80'E 40°E 30°N 20°N 10°N 40°E FORECAST 06-06-2013: DAY 5-11 40°E 30°N 20°N 10°N FORECAST 23-05-2013: DAY 19-25 40°E 30"N 20°N 10°N

Analysis and ECMWF EPS-Monthly Forecasting System Precipitation anomaly

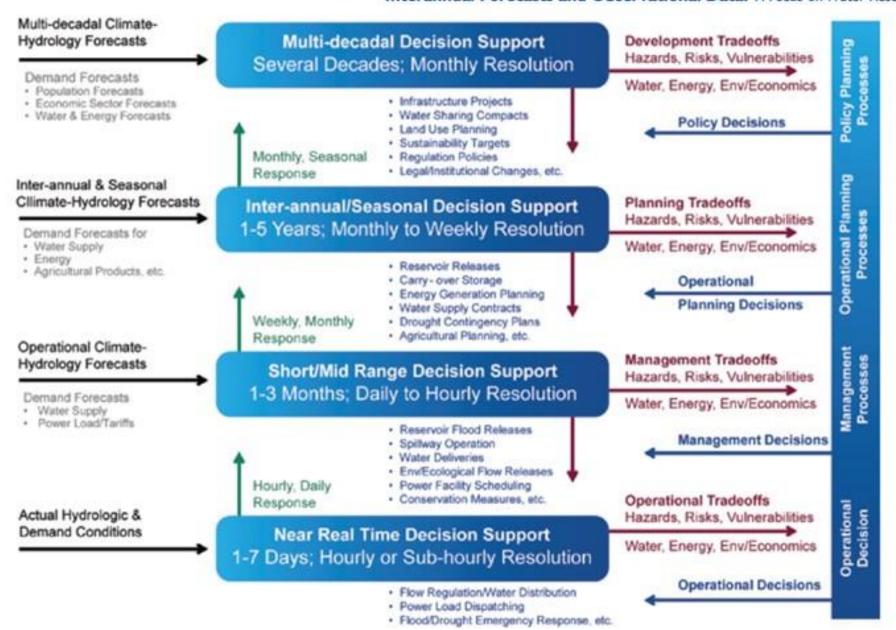
Verification period: 10-06-2013/TO/16-06-2013

ensemble size = 51 ,climate size = 100 Shaded areas significant at 10% level Contours at 1% level



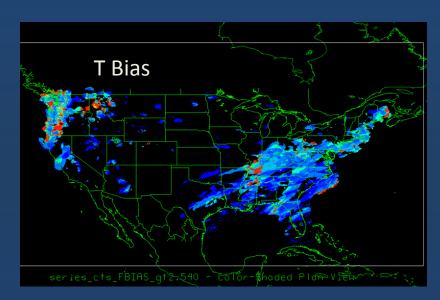
But is seamless verification practical?

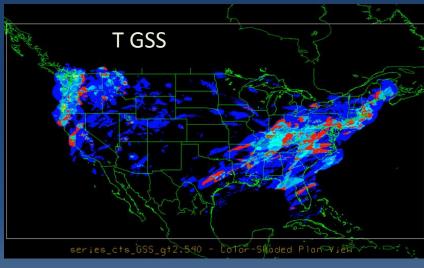
- Observation and sample size limitations increase with increasing forecast length
 - Greater aggregation => reduced specificity
- Users make different kinds of decisions on different time scales
 - Problems are different on different time scales
 - Consequences of decisions change
 - Is it practical to think we can use the same approaches for evaluation?



Weather predictions

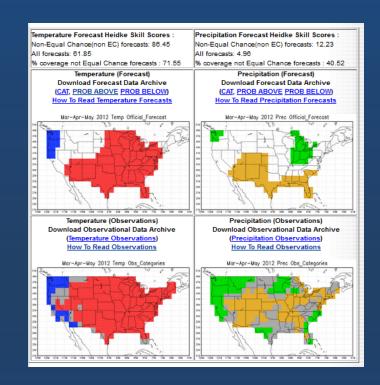
- Smaller scale grids and more frequent forecasts
 - Possible to examine forecasts on local scales (e.g., point verification) and short time scales
 - Possible to link forecast performance to environmental conditions (i.e., stratification)
- Small errors or variations in errors can be important
- Possible to establish statistical significance

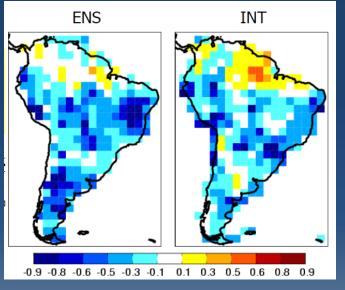




Seasonal predictions

- Typical verification approaches examine "success" of 3-category predictions through standard verification approaches
- May use a gridded approach
- Less opportunity to subset results
- Difficult to do extensive forecast evaluation without hindcasts

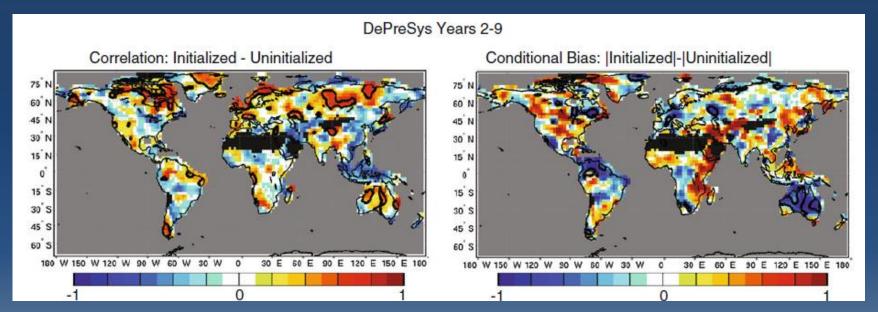




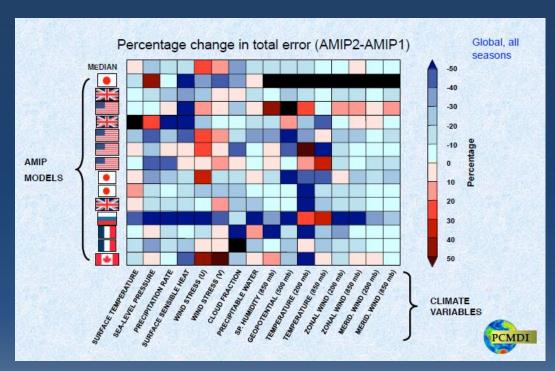
Brier Skill Score Hindcast (1959-2001) evaluation; 1-mo precip forecast (from DEMETER)

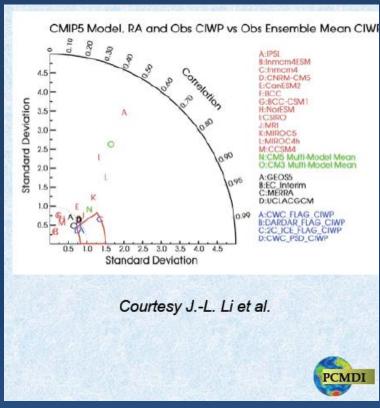
Climate metrics: Interannual to Decadal and longer

- Typically larger gridded regions, standard metrics (e.g., correlation)
- Hindcasts are the focus... Difficult (or impossible) to evaluate actual projections



Climate Model Intercomparisons



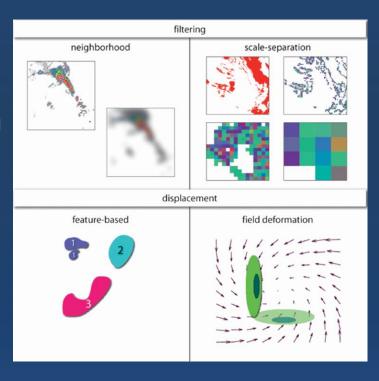


Summary comments

- The options for evaluation of forecasts at ISI time scales are limited by
 - Event frequencies (sample sizes)
 - Spatial and temporal correlations
- Evaluation approaches should be defined by user applications which depend strongly on spatial and temporal scale
- Unique but overlapping verification approaches are required for crossing scales

Summary comments cont.

- Even with small sample sizes longer-range forecasts are amenable to application of new or alternative verification approaches
 - Spatial methods (can be applied to single pairs of images)
 - Extreme dependency scores may give more information about performance for extremes
 - User-relevant metrics



Things to think about

- What are the most meaningful metrics for different spatial and temporal scales?
- How can we provide representative information that represents the needs of a relevant range of users?

(Ignorance score?)

 How do we balance the needs for information with the smaller sample sizes available at longer lead times?

WMO Sub-seasonal to Seasonal (S2S) Project

- Under WMO World Weather Research Project (WWRP)
- Focus: Improve S2S predictions
 - Sharing operational forecasts and observations
 - Examining in detail specific extreme events
- Verification
 - Establishment of recommended metrics on S2S time scale
 - More specificity than Seasonal or Climate time scales
 - Less specificity than Weather time scale
- Workshop at NCEP in Feb 2014